

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-160432

(43)Date of publication of application : 23.06.1995

(51)Int.Cl. G06F 3/06
G06F 11/16
G06F 12/08

(21)Application number : 05-308995

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 09.12.1993

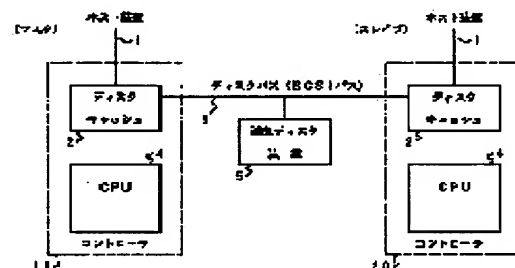
(72)Inventor : ISHII TAKASHI

(54) DUPLEX SYSTEM FOR MAGNETIC DISK CONTROLLER

(57)Abstract:

PURPOSE: To construct the duplex system for the magnetic disk controller which has an environment wherein the high reliability of a magnetic disk device having a write-back type disk cache is easily realized by copying cache data of a controller of one system to the other system at normal time.

CONSTITUTION: The system is provided with a data transfer path 3 which copies the contents of the cache data and control information on the storage destination address, etc., of the data from the controller 10 that normally responds to a command from a host device to the controller 20 normally in a stand-by state, and, the same contents are made normally present in the caches 2 and 2 of the controllers of both the systems, but if the controller 10 gets out of order, the cache data are written to the magnetic disk device 5 from the controller 20.



LEGAL STATUS

[Date of request for examination] 09.09.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3122295

[Date of registration] 20.10.2000

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998.2003 Japan Patent Office

**FULLY ENGLISH TRANSLATION OF JAPANESE LAID-OPEN PATENT HEI
7-160432**

[Claim(s)]

5 [Claim 1] A disk cache system in which plural disk controller
having cache share a magnetic disk unit and for returning
the status of a processing end to a host equipment at the
time when the disk controller receives data generated by
the host, then writing data in a magnetic disk unit, has
10 the path which communicates among both the above-mentioned
magnetic disk controllers as an access path to a share
magnetic disk unit,

Wherein the system copies cache data which one magnetic
~~disk controller has to the cache of another magnetic disk~~
15 controller via the path, recognizes data in both magnetic
disk controllers, and writes the data to a magnetic disk
unit via the magnetic disk controller which is directed
by the host equipment.

[Claim 2] The duplex method of the magnetic disk controller
20 according to claim 1, wherein the magnetic disk controller
in an operating condition always copies and sends cache
data to the magnetic disk controller in a standby state,
and the cache data is written to a magnetic disk unit through
the magnetic disk controller in a standby state at the time
25 of failure of the magnetic disk controller in an operating
condition.

[Claim 3] The duplex method of the magnetic disk controller

according to claim 1, wherein plural said magnetic disk controllers is build in CPU, respectively,

And wherein this CPU issues the command including a write address to another magnetic disk controller, writes
5 also in the cache of an other system through an access path at the same time of writing the data obtained from host equipment in the cache of itself, receives a command from another magnetic disk controller after returning the status to the host equipment, acquires the directory information
10 on the data which another magnetic disk controller wrote in to the magnetic disk unit, and prepares for update of a directory of itself.

[Detailed Description of the Invention]

[0001]

15 [Industrial Application] This invention relates to a duplex method of a magnetic disk controller which is favorable to be used in disk cache system in which plural disk controller having cache share a magnetic disk unit and is adopted a write back method for returning the status of
20 a processing end to a host equipment at the time when the disk controller receives data generated by the host, then writing data in a magnetic disk unit.

[0002]

[Description of the Prior Art] A magnetic disk unit is used
25 abundantly as mass external file apparatus in a field of not only a large-sized computer but also a distributed-processing computer and recently a personal

computer, and has been miniaturized with a high density every year.

[0003] In order to secure the reliability of this system, the system which doubled the magnetic disk controller which
5 controls a magnetic disk unit is built. In the conventional system which doubled such a magnetic disk controller, the access path to a magnetic disk unit is provided to both magnetic disk controllers, and when one magnetic disk controller broke down, it is adopted to maintain the access
10 to a magnetic disk unit from the other magnetic disk controller.

[0004] In the latest magnetic disk controller, a disk cache of write back method, that the magnetic disk controller receives first the write data from a host (equipment by
15 the side of a host), reports the status of a processing end to the host at the time, and then writes data to the magnetic disk unit, is used often. This disk cache serves as an indispensable support item, in order to carry out rapid access of the magnetic disk unit.

20 [0005] When using the disk cache of the write back method mentioned above, it is necessary to write the data which remained in the cache memory to a magnetic disk unit after failure of a magnetic disk controller. Therefore, in the conventional duplex method which prepares two magnetic disk
25 controllers having a cache memory simply, it is impossible to write data which is not written to the magnetic disk unit.

[0006] Then, although there was also a method which separates a cache memory from a magnetic disk controller, and provides the access path from both magnetic disk controllers in order to avoid this problem, when such a method was used, the special cache device needed to be connected in the case of not doubling a magnetic disk controller, therefore there was a problem from which the hardware structure of a system becomes complicated and becomes expensive.

[0007]

10 [Problem(s) to be Solved by the Invention] As described above, when using the disk cache of a write back method, it is necessary to write the data which remained in the cache memory to the magnetic disk unit after magnetic disk controller failure. ~~Therefore, it is impossible to write~~
15 data which is not written to the magnetic disk unit in the conventional duplex method which prepares two magnetic disk controllers having a cache memory simply. Then, although the method which separates a cache memory from a magnetic disk controller and provides the access path from both
20 magnetic disk controllers, was also considered, this method needed to connect the special cache device, when not doubling a magnetic disk controller, and had the problem that the hardware structure of a system became complicated and became expensive.

25 [0008] This invention was made in view of the above-mentioned actual condition, and the object of this invention is to provide the duplex method of a magnetic disk controller

with the environment where high-reliability of the magnetic disk unit having the disk cache of a write back method is realized easily by usually copying cache data of one magnetic disk controller to another magnetic disk controller.

5 [0009]

[Means for Solving the Problem] This invention is characterized that in a disk cache system in which plural disk controller having cache share a magnetic disk unit and for returning the status of a processing end to a host
10 equipment at the time when the disk controller receives data generated by the host, then writing data in a magnetic disk unit, has the path which communicates among both the above-mentioned magnetic disk controllers as an access path
~~to a share magnetic disk unit, the system copies cache data~~
15 which one magnetic disk controller has to the cache of another magnetic disk controller via the path, recognizes data in both magnetic disk controllers, and writes the data to a magnetic disk unit via the magnetic disk controller which is directed by the host equipment. Moreover, this
20 invention is characterized that the magnetic disk controller in an operating condition always reproduces and sends cache data to the magnetic disk controller in a standby state, and it is performed to write the cache data to a magnetic disk unit through the magnetic disk controller
25 in a standby state at the time of failure of the magnetic disk controller in an operating condition.

[0010]

[Function] In this invention, data transfer path and processing procedure are provided in order to copy control information such as the contents of the cache data and storage destination address of the data from the magnetic disk controller system which usually answers a command from host equipment to the magnetic disk controller system which works when the magnetic disk controller system breaks down and which is usually in a standby state. Usually it is operated that the same content exists in the cache of the magnetic disk controller of a both system, and at the time of failure of one magnetic disk controller, it is adopted writing method which writes cache data to a magnetic disk from another magnetic disk controller which remains. By this method, duplex of the magnetic disk controller which has the disk cache of a write back method is realized.

[0011] In order to realize the above-mentioned processing, CPU is built in each magnetic disk controller. By control of this CPU, as shown in Fig.2, CPU issues the command including the write address to the magnetic disk controller of another side, writes also in the cache of the other system prepared in the magnetic disk controller of another side through the above-mentioned access path while writing the data obtained from host equipment in the cache of itself. After returning the status to host equipment, a command is received from the magnetic disk controller of another side, the directory information on the data which the magnetic disk controller of another side wrote in to the

magnetic disk unit is acquired to prepare to update a directory of itself.

[0012] By this, when not need reliability so much, doubling of a magnetic disk controller does not make just like a conventional duplex method of the magnetic disk controller which does not use the disk cache of a write back method, when need a high reliability, by adding only a magnetic disk controller to system, a high reliability system is constructed easily.

10 [0013]

[Embodiments] Hereafter, one embodiment of this invention is explained using a drawing. Fig.1 shows a block diagram showing one embodiment of this invention. In drawing, a sign 1 is a host bus used in order to connect with the host

15 equipment which is not illustrated, and the host equipment performs data transfer between the disk caches 2 in a magnetic disk controller through this bus 1.

[0014] A sign 2 is a disk cache in a magnetic disk controller, and performs data transfer with the host bus 1 and the disk bus 3 mentioned later. A sign 3 is a disk bus and can perform communication between magnetic disk controllers via for example, a SCSI bus with transmission and reception of the command data to a magnetic disk unit 5.

[0015] A sign 4 is a microprocessor (CPU) in a magnetic disk controller, and performs processing control of the magnetic whole disk controller. A sign 5 is a magnetic disk unit and is used with a share by two or more magnetic disk

controllers 10 and 20 through the disk bus 3. Here, the magnetic disk controller 10 is called the magnetic disk controller by the side of a master, and the magnetic disk controller 20 is called the magnetic disk controller by the side of a slave.

[0016] Fig.2 is a flowchart in which operation of the embodiment of this invention is shown. Fig.2(a) shows the operations sequence of write command processing of the microprocessor 4 in the magnetic disk controller 10 used as a master, and Fig.2(b) shows the operations sequence of write command processing of the microprocessor 4 in the magnetic disk controller 20 used as a slave.

[0017] In Fig.2, sign Sa is a step to which the magnetic disk controller 10 publishes a command to the magnetic disk controller 20 of another system, and this step notifies control information including the storing place address of data to the magnetic disk controller 20 of another system, and makes data transfer prepare.

[0018] Sign Sb is a step in which the magnetic disk controller 10 performs data transfer. In this step, the data obtained through the host bus 1 is written in the disk cache 2 (the magnetic disk controller 10 has) of itself and also the data is written in the disk cache (the magnetic disk controller 20 has) 2 of another system via the disk bus 3.

[0019] Sign Sc is a step in which the magnetic disk controller 10 receives the command issued from the magnetic disk

controller 20 of another system. In this step, the controller
10 receives the directory information on the data which
the magnetic disk controller 20 of an other system wrote
in the magnetic disk unit 5, and prepares renewal of a
5 directory by the magnetic disk controller 10 of itself.
[0020] Sign Sd is a step to which the magnetic disk controller
20 performs disk writing. And the magnetic disk controller
20 in a standby state usually takes charge of processing
in the sense of a load distribution. Sign Sf is a step to
10 which the magnetic disk controller 20 publishes the updating
demand of a directory. In this step, the directory breathed
out from the disk cache 2 is notified to the magnetic disk
controller 10 of another system to take the adjustment of
~~the content of a disk cache of a both system~~

15 [0021] Hereafter, operation of the embodiment of this
invention is explained. As shown in the Fig.1, in this
invention, the magnetic disk controller of two systems is
connected through the disk bus in which communication
between magnetic disk controllers is possible. Usually,
20 the magnetic disk controller 10 shown as a master receives
the command from the host which does not illustrate, and
performs processing according to instruction of the host
equipment. However, for only data write request to a disk
unit 5, operation shown a flowchart in Fig.2 is performed.
25 That is, when data are written in the disk caches built-in
in the magnetic disk controller used as itself and a slave,
processing is finished, and the write-in operation to a

disk unit 5 itself is left to a slave.

[0022] Although the magnetic disk controller 20 shown as a slave is a magnetic disk controller for standby of the sake at the time of failure of the magnetic disk controller 10 used as a master, as shown in Fig.2, it operates as a write-back cache controller which performs processing which writes cache data in a magnetic disk unit 5 at the time of standby.

[0023] When the magnetic disk controller 10 shown as a master breaks down, the magnetic disk controller 20 shown as a slave receives the command from host equipment, and comes to perform whole command processing to a magnetic disk unit 5.

~~[0024] In this time, since write data to which the magnetic~~
disk controller 10 which already serves as a master was processing exists in the disk cache 2 of itself, the controller 10 can also write the content in a magnetic disk unit 5 satisfactory.

[0025] On the other hand, when the magnetic disk controller 20 shown as a slave breaks down, the magnetic disk controller 10 used as a master performs write-in processing to a magnetic disk unit 5 instead of a slave.

[0026] At this time, since required data are in the disk cache 2 of itself, continuation of operation can be performed satisfactory. When the magnetic disk controller of a piece system breaks down, a report to that effect is notified to the host equipment. At this time, it is left to judgment

of host equipment whether cache operation of a write back method is made to perform only by the magnetic disk controller of one system. Moreover, write buffer cache operation is possible for the magnetic disk controller alone
5 shown in Fig.1 when a magnetic disk controller was not doubled. Therefore, in the system which does not need reliability so much, write-back cache operation is made to perform by a magnetic disk controller alone, if need, it can improve reliability only by adding a magnetic disk
10 controller.

[0027]

[Effect of the Invention] According to this invention, like explanation above by usually copying cache data of one magnetic disk controller to another magnetic disk
15 controller, it is easily realized to duplex the controller. And, by this, when not need reliability so much, doubling of a magnetic disk controller does not make just like a conventional duplex method of the magnetic disk controller which does not use the disk cache of a write back method,
20 when need a high reliability, by adding only a magnetic disk controller to system, a high reliability system is constructed easily.

[Brief Description of the Drawings]

[Fig.1] The block diagram showing the composition of the
25 embodiment of this invention.

[Fig.2] The flowchart in which operation of the embodiment of this invention is shown.

[Description of Notations]

1 -- A host bus, 2--A disk cache, 3-- A disk bus (SCSI bus),
4 -- A microprocessor (CPU), 5-- A magnetic disk unit, 10,
20 -- Magnetic disk controller.

5

FIG.1:

1 -- A host bus,
2--A disk cache,
3-- A disk bus (SCSI bus),
5 4 -- A microprocessor (CPU),
5-- A magnetic disk unit,
10, 20 -- Magnetic disk controller.

FIG.2(a)

10 Master start
Receive command
Calculate cache ADR
Sa--Issue command to another system
Sb--Transfer data
15 Check status of another system
Report completion
Sc--Receive command from another system
Update DIR
END
20

FIG.2(b)

Slave start
Receive command from another system
Transfer data
25 Report completion
Sd--Write disk
Se--DIR change command

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-160432

(43)公開日 平成7年(1995)6月23日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	3 0 4 E			
11/16	3 1 0 F			
12/08	3 2 0	7608-5B		

審査請求 未請求 請求項の数3 O L (全 5 頁)

(21)出願番号 特願平5-308995

(22)出願日 平成5年(1993)12月9日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 石井 隆

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

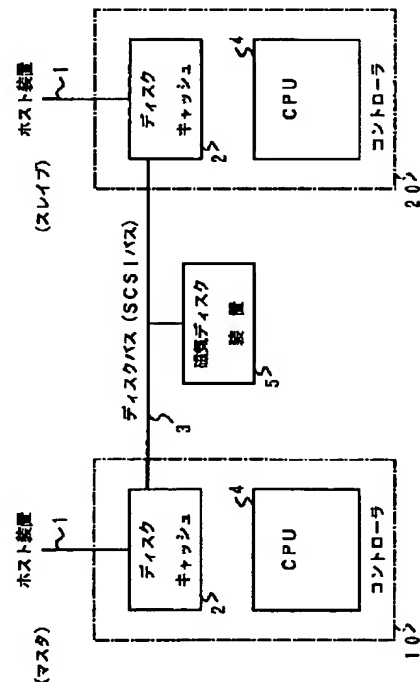
(74)代理人 弁理士 鈴江 武彦

(54)【発明の名称】 磁気ディスクコントローラの二重化方式

(57)【要約】

【目的】この発明は、通常時に片系コントローラのキャッシュデータを他系に複製しておくことにより、ライトバック方式のディスクキャッシュを有する磁気ディスク装置の高信頼性を容易に実現する環境を持った磁気ディスクコントローラの二重化方式を構築することを主な特徴とする。

【構成】通常時にホスト装置からのコマンドに応答するコントローラ10から、そのコントローラが故障した時に稼働するため、通常、スタンバイ状態にあるコントローラ20へキャッシュデータの内容とデータの格納先アドレス等の制御情報を複写するデータ転送経路3を設け、通常時は両系のコントローラのキャッシュ2、2に同一の内容が存在するようにして稼働させ、コントローラ10が故障時はコントローラ20からキャッシュデータを磁気ディスク装置5へ書き込むことを特徴とする。



【特許請求の範囲】

【請求項 1】 キャッシュをもつ複数の磁気ディスクコントローラが磁気ディスク装置を共有し、上記磁気ディスクコントローラがホスト装置より生成されるデータを受信すると、ホスト装置に対しその時点で処理終了の旨のステータスを返し、その後に磁気ディスク装置にデータの書き込みを行なうディスクキャッシュシステムに於いて、上記磁気ディスクコントローラ相互の間で通信を行うパスを、共有磁気ディスク装置に対するアクセスパスとして備え、このパスを介して一方の磁気ディスクコントローラが持つキャッシュデータを他方のキャッシュに複製し、双方の磁気ディスクコントローラにてデータの認知を行ない、ホスト装置によって指示される磁気ディスクコントローラ経由で磁気ディスク装置に対してデータの書き込みを行うことを特徴とする磁気ディスクコントローラの二重化方式。

【請求項 2】 稼働状態にある磁気ディスクコントローラが、待機状態にある磁気ディスクコントローラに対し常にキャッシュデータを複製して送付し、故障時に、待機状態にある磁気ディスクコントローラを介してキャッシュデータを磁気ディスク装置に対し書き込むことを特徴とする請求項 1 記載の磁気ディスクコントローラの二重化方式。

【請求項 3】 複数の磁気ディスクコントローラは、それぞれ CPU を内蔵し、この CPU は、他方の磁気ディスクコントローラに対して書き込みアドレスを含むコマンドを発行し、ホスト装置から得られるデータを自系のキャッシュへ書き込むと同時にアクセスパスを介して他系のキャッシュへも書き込み、ホスト装置に対してステータスを返した後、他方の磁気ディスクコントローラからコマンドを受信し、他方の磁気ディスクコントローラが磁気ディスク装置に対し書き込んだデータのディレクトリ情報を得、自系のディレクトリ更新に備えることを特徴とする請求項 1 記載の磁気ディスクコントローラの二重化方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、磁気ディスク装置を複数の磁気ディスクコントローラで共有し、ホスト装置によって生成されるデータを上記磁気ディスクコントローラが受信し、ホスト装置に対してその時点で処理終了の旨のステータスを返し、その後に磁気ディスク装置に対するデータの書き込みを行うライトバック方式を採用したディスクキャッシュシステムに用いて好適な磁気ディスクコントローラの二重化方式に関する。

【0002】

【従来の技術】 磁気ディスク装置は大型コンピュータのみならず、分散処理コンピュータ、最近ではパーソナルコンピュータの分野に於いても大容量外部ファイル装置として多用され、年々、小型化、高密度化されてきてい

る。

【0003】 この種システムの信頼性を確保するために、磁気ディスク装置のコントロールを行う磁気ディスクコントローラを二重化したシステムが構築されている。従来、このような磁気ディスクコントローラを二重化したシステムに於いては、磁気ディスク装置に対するアクセスパスを双方の磁気ディスクコントローラに対して持たせ、片系の磁気ディスクコントローラが故障した場合に、他系の磁気ディスクコントローラから磁気ディスク装置へのアクセスが維持できる方法が採られていた。

【0004】 最近の磁気ディスクコントローラでは、ホスト（ホスト側の装置）からの書き込みデータを磁気ディスクコントローラが先ず受取り、ホストに対しその時点で処理終了の旨のステータスを報告し、その後に磁気ディスク装置へのデータ書き込みを行うライトバック方式のディスクキャッシュが用いられることが多くなった。このディスクキャッシュは磁気ディスク装置を高速アクセスするために必須のサポート項目となっている。

【0005】 上述したライトバック方式のディスクキャッシュを用いる場合には、磁気ディスクコントローラの故障後に於いて、キャッシュメモリに残ったデータを磁気ディスク装置へ書き込む必要がある。従って、単純にキャッシュメモリを持つ磁気ディスクコントローラを 2 個用意する従来の二重化方式では、磁気ディスク装置に対しての末書き込みデータの書き込みが不可能である。

【0006】 そこで、この不都合を回避するため、キャッシュメモリを磁気ディスクコントローラから切り放し、双方の磁気ディスクコントローラからのアクセスパスを持たせる方式もあるが、このような方式を用いる場合には磁気ディスクコントローラを二重化しない場合にも特殊なキャッシュデバイスを接続する必要があり、システムのハードウェア構造が複雑となり高価となる問題があった。

【0007】

【発明が解決しようとする課題】 上記したように、ライトバック方式のディスクキャッシュを用いる場合には、磁気ディスクコントローラ故障後にキャッシュメモリに残ったデータを磁気ディスク装置へ書き込む必要があり、従って、単純にキャッシュメモリを持つ磁気ディスクコントローラを 2 個用意する従来の二重化方式では、磁気ディスク装置への末書き込みデータの書き込みが不可能である。そこで、キャッシュメモリを磁気ディスクコントローラから切り放し、双方の磁気ディスクコントローラからのアクセスパスを持たせる方式も考えられたが、この方式は、磁気ディスクコントローラを二重化しない場合にも特殊なキャッシュデバイスを接続する必要があり、システムのハードウェア構造が複雑となり高価になるという問題があった。

【0008】 本発明は上記実情に鑑みなされたもので、

通常時に片系磁気ディスクコントローラのキャッシュデータを他系に複製しておくことにより、ライトバック方式のディスクキャッシュを有する磁気ディスク装置の高信頼性を容易に実現する環境を持った磁気ディスクコントローラの二重化方式を提供することを目的とする。

【0009】

【課題を解決するための手段】この発明は、磁気ディスク装置を複数の磁気ディスクコントローラで共有し、ホスト装置によって生成されるデータを上記磁気ディスクコントローラが受信し、ホスト装置に対してその時点で処理終了の旨のステータスを返し、その後磁気ディスク装置に対するデータの書き込みを行うディスクキャッシュシステムに於いて、双方の磁気ディスクコントローラ間で通信を行うバスを、共有する磁気ディスク装置に対するアクセスバスとして備え、このバスを介して一方の磁気ディスクコントローラが持つキャッシュデータを他方のキャッシュに複製し、双方の磁気ディスクコントローラにてデータの認知を行ない、ホスト装置によって指示される磁気ディスクコントローラ経由で磁気ディスク装置に対してデータの書き込みを行うことを特徴とする。又、稼働状態にある磁気ディスクコントローラは、待機状態にある磁気ディスクコントローラに対し常にキャッシュデータを複製して送付し、稼働状態にある磁気ディスクコントローラの故障時に、待機状態にある磁気ディスクコントローラを介してキャッシュデータを磁気ディスク装置に対し書き込むことを特徴とする。

【0010】

【作用】この発明は、通常時にホスト装置からのコマンドに応答する磁気ディスクコントローラ系から、その磁気ディスクコントローラ系が故障した時に稼働する、通常スタンバイ状態にある磁気ディスクコントローラ系へ、キャッシュデータの内容とデータの格納先アドレス等の制御情報を複写するデータ転送経路と処理手順を設け、通常時は、両系の磁気ディスクコントローラのキャッシュに同一の内容が存在するようにして稼働させ、片系磁気ディスクコントローラの故障時には、残る別系の磁気ディスクコントローラからキャッシュデータを磁気ディスクへ書き込む方式を採用することにより、ライトバック方式のディスクキャッシュを有する磁気ディスクコントローラの二重化を実現する。

【0011】そこで上記した処理を実現するため、各磁気ディスクコントローラにCPUを内蔵し、このCPUの制御により、図2に示すように、他方の磁気ディスクコントローラに対して書き込みアドレスを含むコマンドを発行し、ホスト装置から得られるデータを自系のキャッシュへ書き込むと同時に上記アクセスバスを介して他方の磁気ディスクコントローラに設けた他系のキャッシュへも書き込み、ホスト装置に対してステータスを返した後、他方の磁気ディスクコントローラからコマンドを受信し、他方の磁気ディスクコントローラが磁気ディス

ク装置に対し書き込んだデータのディレクトリ情報を得、自系のディレクトリ更新に備える。

【0012】このことにより、ライトバック方式のディスクキャッシュを用いない従来の磁気ディスクコントローラの二重化と同様、さほど信頼性を必要としない場合には磁気ディスクコントローラを二重化せず、高信頼性が必要な場合には磁気ディスクコントローラのみを追加することにより、容易に高信頼システムを構築できる。

【0013】

【実施例】以下、図面を使用してこの発明の一実施例について説明する。図1はこの発明の一実施例を示すブロック図である。図に於いて、符号1は図示しないホスト装置と接続するために用いられるホストバスであり、ホスト装置はこのバス1を介して磁気ディスクコントローラ内のディスクキャッシュ2との間のデータ転送を行う。

【0014】符号2は磁気ディスクコントローラ内のディスクキャッシュであり、ホストバス1、及び後述するディスクバス3とのデータ転送を行う。符号3はディスクバスであり、磁気ディスク装置5へのコマンド・データの送受とともに、例えばSCSIバス経由での磁気ディスクコントローラ間通信が行える。

【0015】符号4は磁気ディスクコントローラ内のマイクロプロセッサ(CPU)であり、磁気ディスクコントローラの全体の処理制御を行う。符号5は磁気ディスク装置であり、ディスクバス3を介して複数の磁気ディスクコントローラ10、20により共有使用される。ここでは、磁気ディスクコントローラ10をマスタ側の磁気ディスクコントローラ、磁気ディスクコントローラ20をスレーブ側の磁気ディスクコントローラと称す。

【0016】図2はこの発明の実施例の動作を示すフローチャートである。図2(a)はマスタとなる磁気ディスクコントローラ10内のマイクロプロセッサ4のライトコマンド処理の動作手順を示し、同図(b)はスレーブとなる磁気ディスクコントローラ20内のマイクロプロセッサ4のライトコマンド処理の動作手順を示している。

【0017】図2に於いて、符号Saは磁気ディスクコントローラ10が他系の磁気ディスクコントローラ20へコマンドを発行するステップであり、データの格納先アドレスを含む制御情報を他系の磁気ディスクコントローラ20に通知し、データ転送の準備を行わせる。

【0018】符号Sbは磁気ディスクコントローラ10がデータ転送を行うステップであり、ここでホストバス1を介して得られるデータを自系の(磁気ディスクコントローラ10がもつ)ディスクキャッシュ2へ書き込むと同時に、ディスクバス3を経由して他系の(磁気ディスクコントローラ20がもつ)ディスクキャッシュ2へ書き込む。

【0019】符号Scは磁気ディスクコントローラ10

が他系の磁気ディスクコントローラ 20 から発せられるコマンドを受け取るステップであり、他系の磁気ディスクコントローラ 20 が磁気ディスク装置 5 へ書き込んだデータのディレクトリ情報を受取り、自系の磁気ディスクコントローラ 10 によるディレクトリ更新に備える。

【0020】符号 S d は磁気ディスクコントローラ 20 がディスク書き込みを行うステップであり、通常、スタンバイ状態にある磁気ディスクコントローラ 20 が負荷分散の意味で処理を受け持つ。符号 S f は磁気ディスクコントローラ 20 がディレクトリ更新要求を発行するステップであり、ディスクキャッシュ 2 から吐き出されたディレクトリを他系の磁気ディスクコントローラ 10 に通知して両系のディスクキャッシュ内容の整合性をとる。

【0021】以下、この発明の実施例の動作について説明する。第 1 図に示したように、本発明では磁気ディスクコントローラ間通信が可能なディスクバスを介して 2 系の磁気ディスクコントローラが接続される。通常はマスタとして示した磁気ディスクコントローラ 10 が図示しないホスト装置からのコマンドを受取り、ホスト装置の指示に従った処理を行う。但し、ディスク装置 5 へのデータ書き込み要求に限り、図 2 にフローチャートで示す動作を実行する。即ち、自系及びスレーブとなる磁気ディスクコントローラに内蔵のディスクキャッシュにデータを書き込んだ時点で処理を終え、ディスク装置 5 への書き込み動作自体はスレーブに任せる。

【0022】スレーブとして示した磁気ディスクコントローラ 20 は、マスタとなる磁気ディスクコントローラ 10 の故障時のための待機用磁気ディスクコントローラであるが、図 2 に示したように、待機時にはキャッシュデータを磁気ディスク装置 5 へデータを書き込む処理を行うライトバックキャッシュコントローラとして動作する。

【0023】マスタとして示した磁気ディスクコントローラ 10 が故障した場合に、スレーブとして示した磁気ディスクコントローラ 20 はホスト装置からのコマンドを受付け、磁気ディスク装置 5 へのコマンド処理全般を行うようになる。

【0024】このとき、既にマスタとなる磁気ディスクコントローラ 10 が処理していた書き込みデータは自系

ディスクキャッシュ 2 内に存在するため、問題なく、その内容も磁気ディスク装置 5 へ書き込める。

【0025】一方、スレーブとして示した磁気ディスクコントローラ 20 が故障した場合は、マスタとなる磁気ディスクコントローラ 10 が磁気ディスク装置 5 への書き込み処理をスレーブに代わって行う。

【0026】このとき、必要なデータはやはり自系ディスクキャッシュ 2 内にあるため、動作の継続は問題なく行える。片系の磁気ディスクコントローラが故障した場合にはその旨の報告はホスト装置に通知される。このとき、片系磁気ディスクコントローラのみでラインバック方式のキャッシュ動作を行わせるか否かはホスト装置の判断に委ねられる。また、磁気ディスクコントローラの二重化を行わない場合にも、図 1 に示した磁気ディスクコントローラは単体でライトバックファキャッシュ動作が可能であり、信頼性をさほど必要としないシステムでは磁気ディスクコントローラ単体でライトバックキャッシュ動作を行わせ、必要となった場合には磁気ディスクコントローラを追加するのみで信頼性を向上できる。

【0027】

【発明の効果】以上説明のように、本発明によれば、通常時に片系磁気ディスクコントローラのキャッシュデータを他系磁気ディスクコントローラのディスクキャッシュに複製しておくことにより、容易にコントロールの二重化を実現でき、ライトバック方式のディスクキャッシュを用いない従来の磁気ディスクコントローラの二重化と同様、さほど信頼性を必要としない場合には磁気ディスクコントローラを二重化せず、信頼性が必要な場合にのみ磁気ディスクコントローラを追加することにより、容易に高信頼システムを構築できる。

【図面の簡単な説明】

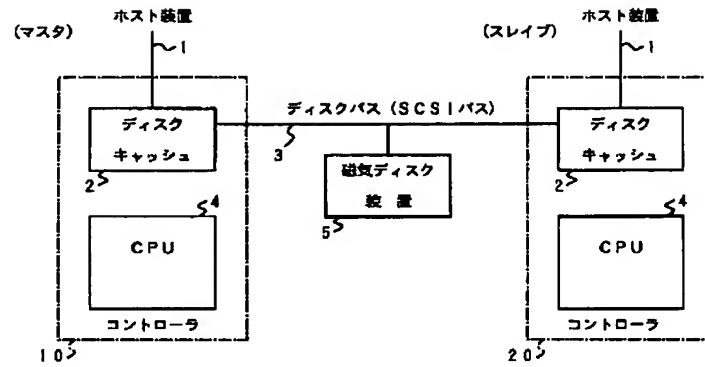
【図 1】この発明の実施例の構成を示すブロック図。

【図 2】この発明の実施例の動作を示すフローチャート。

【符号の説明】

1…ホストバス、2…ディスクキャッシュ、3…ディスクバス（SCSI バス）、4…マイクロプロセッサ（CPU）、5…磁気ディスク装置、10、20…磁気ディスクコントローラ。

【図 1】



【図 2】

